

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12Q 1/68		A1	(11) International Publication Number: WO 00/43538
			(43) International Publication Date: 27 July 2000 (27.07.00)
(21) International Application Number: PCT/CA00/00047		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 19 January 2000 (19.01.00)			
(30) Priority Data: 2,259,745 19 January 1999 (19.01.99) CA			
(71) Applicant (for all designated States except US): UNIVERSITE DE MONTREAL [CA/CA]; Postal Code 6128, Station A, Montreal, Quebec H3C 3J7 (CA).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): PAQUIN, Bruno [CA/CA]; 289 Randill, Châteauguay, Quebec J6J 2P4 (CA). BRUKNER, Ivan [CA/CA]; 1882 Sherbrooke East #2, Montreal, Quebec H2K 1B5 (CA). TREMBLAY, Guy [CA/CA]; 3341 Maréchal #4, Montreal, Quebec H3T 1M8 (CA).			
(74) Agents: DUBUC, Jean, H. et al.; Goudreau Gage Dubuc & Martineau Walker, The Stock Exchange Tower, Suite 3400, 800 Place Victoria, P.O. Box 242, Montreal, Quebec H4Z 1E9 (CA).			
(54) Title: PROCESS FOR THE GENERATION OF OLIGONUCLEOTIDE LIBRARIES (OLs) REPRESENTATIVE OF GENOMES OR EXPRESSED mRNAs (cDNAs) AND USES THEREOF			
(57) Abstract			
<p>A process for the generation of oligonucleotide libraries representative of a given template is described. Starting from a random pool of oligonucleotides, the process selects only those which hybridize to the template nucleic acid. This selection yields a highly specific library that represents an oligo-image of the chosen template. The novel quality of this approach is the generation of amplifiable oligonucleotide probes that are of uniform length, free of repetitive sequence motifs and easily subjected to differential selection. This technique is used to produce different oligonucleotide libraries (OLs) and shows that these OLs do not cross-hybridize. Differential selection of these OLs produces oligonucleotides that can be used in the identification, characterization and isolation of nucleic acids.</p>			
<p>Membrane-bound denatured target DNA</p> <p>Random 20-mer core with left and right blockers</p> <p>ds ss ds</p> <p>Preparative hybridization</p> <p>1. Wash unbound OL 2. Elute bound OL 3. PCR amplify bound OL</p> <p>ds ds ds</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AI	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauntania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakistan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

TITLE OF THE INVENTION

Process for the Generation of Oligonucleotide Libraries (OLs)
Representative of Genomes or Expressed mRNAs (cDNAs) and Uses
5 Thereof

FIELD OF THE INVENTION

The present invention relates to a process for the generation
10 oligonucleotide libraries (OLs) representative of genomes or expressed
mRNAs (cDNAs) and to the uses thereof. In particular, the present
invention relates to a process for the generation of oligonucleotide
libraries comprising oligonucleotides of uniform length. The present
invention further relates to the uses of these OLs in numerous
15 biotechnological applications, including the identification and/or
characterization of biological materials, clinical diagnosis (DNA/RNA
level), preparative extraction of specific mRNA (and genes) and genomic
research/mapping.

20 BACKGROUND OF THE INVENTION

The generation of genomic DNA libraries, or cDNA libraries and the
maintenance, and handling of these libraries are critical procedures in the
field of genomics and/or biotechnology. In classical libraries the relevant
25 segments of DNA are cloned into vectors, which are maintained and
propagated in particular biological systems (*in vivo*). Alternatively, libraries
(*in vitro*) can be directly constructed from genomic DNA or cDNA. They
contain linkers at the 5' and 3' ends of the DNA which allow PCR
amplification of the library. The information stored in these libraries

contains repetitive sequence elements that originated from repetitive DNA, or high copy mRNAs. This results in a significant redundancy, which can complicate the use and the outcome of using classical libraries.

Another important feature which reduces the utility of classical libraries is the heterogeneity in size of the members comprising the library. This limits the usefulness of classical libraries in subtractive hybridization procedures (1-2) which are dependent upon the length, complexity and the redundancy of the libraries, and which therefore are particularly sensitive to the choice of method and the number of cycles performed.

In fact, one must tailor the hybridization conditions to accommodate the heterogeneous length and redundancy of stored information in order to perform subtraction. Thus, the results are more "laboratory-specific" than library-specific.

A number of diagnostic methods that involve nucleic acid hybridization have arisen in recent years. Most of them are designed to provide qualitative information about the presence of a specific sequence motif in a complex analytical mixture of nucleic acids and use a detection system based on PCR and/or DNA chip hybridization technologies (3-7). For both of these technologies, diagnostic oligonucleotides constitute an essential part of the detection system. These oligonucleotides are primarily chosen based on the sequence data of the nucleic acids to be detected. In spite of the power of hybridization to correctly identify a complementary strand, it does face limitations. In fact, the difference in stability between a perfectly matched complement and a complement mismatched at only one base can be as little as 0.5° C (8). This is the fundamental limitation to the power of DNA hybridization for specific identification of a cognate strand. Therefore, the diagnostic power of any chosen oligonucleotide must be validated using an analytical mixture whose sequence context is

not totally known. The problem of adequate probe selection is time and labour-consuming. On the other hand, the growing complexity of detection systems based on oligonucleotide technologies requires a fast selection of a large number of short oligonucleotides.

- 5 Akopyants *et al* (7) performed subtractive hybridization using bacterial DNAs digested by high-frequency restriction enzymes. The use of such restriction enzymes tends to generate DNA fragments having a broadly similar size, about 500 base pairs. However, the uniformity is not rigorous. Moreover, the library created by these restriction fragments still
10 contains a significant number of redundant sequences; consequently, patches of short polymorphism embedded in homologous sequences are going to be missed when such a library is used.

- U.S. Patent No. 5,270,163 (8) teaches a method for the isolation of
15 nucleic acids using high-affinity nucleic acid ligands. This method has been termed the SELEX method (Systematic Evolution of Ligands by Experimental Enrichment) and is based on the use of proteins or small molecules, but not nucleic acids, as targets. The selection of oligonucleotides in the SELEX method relies on the three-dimensional
20 (3D) shape of the oligonucleotides and their fit into the structures of the target molecules. In contrast to this, the selection of oligonucleotides in the present invention is based on hybridization with target nucleic acid.

- Armour *et al* (11) describes the quantitative recovery of amplifiable
25 probes hybridised to an immobilised target. The amplifiable probes consist of PCR or restriction fragments and their technique is meant to assess the copy number of *loci*.

There thus remains a need for oligonucleotide libraries which allow for the use of uniform hybridization conditions to perform selection and/or subtraction while minimizing or eliminating redundant sequences. Advantageously, these libraries can be used in the selection of highly
5 informative and target-specific probe libraries. The present invention seeks to meet these and other needs.

SUMMARY OF THE INVENTION

10 The procedure described herein results in the generation and selection of oligonucleotide probes with a high specificity for a given system. These oligonucleotides cover the entire length of the target DNA, thus increasing detectability which might be lost in classical oligo-detection systems due to secondary DNA structure or DNA deletions present in an analyte
15 mixture. At the same time, they present inexpensive variants of a multiplex oligonucleotide-detection approach, since they are not required to be individually synthesized.

More specifically, in accordance with the present invention, there is
20 provided a process for the generation of oligonucleotide libraries, or OLs. The present invention teaches a process for generating OLs from genomic DNAs and cDNAs, and for performing the subtraction of these libraries.

25 The present invention further teaches OLs which allow the use of hybridization conditions which are controllable and reproducible. In addition, the invention teaches a process for the selection of uniform length OLs which minimizes or eliminates redundant sequences and reduces complexity. The result is the production of highly-informative and
30 target-specific probe libraries.

An object of the present invention is therefore to provide a process for the generation of oligonucleotide libraries comprising OLs of uniform length which are self-amplifiable and easily subjected to subtraction.

5

Another object of the invention is to provide OLs which are compatible with DNA array technology. Indeed, an array of diverse mixtures of oligonucleotides which show differential hybridization patterns could be the best choice for the next generation of DNA diagnostics.

10

Other objects, advantages and features of the present invention will become more apparent upon reading of the following non-restrictive description of preferred embodiments thereof, given by way of example only with reference to the accompanying drawings.

15

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Schematic representation of the experimental procedure for the preparation of OL. Denatured DNA is bound to the membrane and hybridized to the random oligonucleotides library in the presence of

20

blockers. These blocking primers disallow the unspecific hybridization of left and right oligonucleotide arms used for PCR amplification of the OL. (ss - single-stranded DNA, ds - double-stranded DNA).

Hybridization and PCR amplification of OL are described more particular below, in the Experimental Methods.

25

Figure 2: Dot blot hybridization of OL targeted against different genomes. The first row represents the dot blot hybridization of random probes with the specified genomic DNA (adenovirus, pBluescript and lambda). The

last row shows dot blot hybridization of mixed adenovirus and lambda-selected OL. The other rows are analytical dot blot hybridizations of selected OLs with each of the genomes indicated. The procedures of preparative and analytical hybridization are described in the Experimental Methods, below.

Figure 3: Specificity and probe distribution of OL generated from adenoviral genome. (A) The corresponding genome and adenoviral DNA were run on a 1% agarose gel stained with ethidium bromide. The type of restriction enzyme and DNA are indicated on the top of each gel lane. (B) Southern hybridization of the same gel using adenovirus OL as an hybridization probe (see Figure 2, row 2). It should be noted that under the experimental conditions, there was no cross-hybridization with either lambda or human DNA. (C) The same membrane was stripped and rehybridized with a OL directed against a 3648 bp-long restriction fragment. This subset of adenovirus OL was prepared by cutting the membrane corresponding to the 3648 bp band from a similar southern blot and reamplified by PCR as described in the Experimental Methods, below. Thus, it is shown that OL specificity may be enhanced by controlling the choice of targeted DNA fragments in the next round of selection.

Figure 4: The distribution of OL along genomic DNA. The densitometric scan of radioactive signal from OL was integrated over total adenoviral genome (Figure 3, lanes 3 and 5) using Scion Image software (Scion corporation, Frederick, Maryland). The signal intensity of OL probes hybridizing to restriction fragments is linearly proportional to the length of DNA.

Figure 5: Subtractive enrichment of OL. (A) The tester OL is presented by the mixture of two genomes (Adenovirus type 2 and Lambda phage). The driver OL was produced from the Lambda genome only. The single stranded (ss) OL from driver DNA was used to pool out the complementary single stranded mixed tester OL. After removing the subtracted fraction, the remainder of the mixed OL was used as a probe in the analytical hybridization step. The mixed OL probes were analyzed by dot blotting as described (B) before subtractive enrichment and (C) after subtractive enrichment by hybridization to genomic Adenovirus and Lambda DNA.

Figure 6: Relative distribution of 20-mers with the different number of mismatches which do hybridize to targeted DNA. The abscissa shows the number of mismatches present in the 20-mer, while the y-axis illustrates the corresponding relative frequencies. The distribution profile was obtained by calculating the number of combinations for each particular number of mismatches which are thermostable at 52° C. The y-axis was normalized to reflect the relative distribution (%) over the total number of captured oligonucleotides (100%). The majority of n-mers captured after the first round of selection will be 20-mers with less than 6 mismatches. This is described further in the Detailed Description, below.

DETAILED DESCRIPTION

The present invention thus provides a process for the generation of oligonucleotide libraries having the following characteristics:

- 1) A uniform length of about 60 bases, comprising a central segment of about 20 bases randomly varied to represent all possible combinations,

and segments of about 20 bases of a defined sequence flanking the central segment on each side;

2) A uniform number of copies for each sequence motif (consequently, there are no differential hybridization kinetics which could originate from the presence of repetitive DNA); and

3) A melting profile which is characterized by a sharp transition from double stranded to single stranded (or vice versa) oligonucleotides. This is a critical advantage in subtractive hybridization procedures.

10 The use of these OLs enhances the specificity of hybridization to nucleic acids isolated from various sources, thereby allowing for the preparation of oligonucleotide mixtures useful in the detection and quantification of specific nucleic acids or nucleic acid mixtures.

15 In one particular embodiment, the starting pool of oligonucleotides is chemically synthesized and consists of a random region of a fixed length (L), flanked by a constant sequence (primer binding sites, PBS). The random oligonucleotide pool covers n copies ($n=1,2,3,\dots$) of all sequence combinations of length L, i.e. 4^L , which is a total of 10^{12} different sequence motifs for $L=20$ nucleotides. The basic length of oligonucleotides is long enough to generate uniform sequence motifs for a particular biological system. The complexity of the library (10^{12}) overcomes the complexity of the template (which is usually between 10^4 - 10^9). The random pool is then hybridized with a nucleic acid template isolated from any selected source and the unbound oligonucleotides are washed away under stringent conditions. The remaining, template-bound oligonucleotides are then subjected to amplification, using PCR or other methods known to those of skill in the art and using primers complementary to the constant

flanking segments, thus producing a library of oligonucleotides capable of selectively hybridizing to nucleic acid templates.

5 The choice of 20-mer for the length of the oligonucleotide library is not arbitrary but is based on the rationale that the length of the particular sequence motif should be long enough to be unique for even the most complex genomes. The L-mer of length L will be, on average, repeated every 4^L base pairs. The longer the L, the greater the average distance between 2 identical sequence motifs of length L will be.

10 However, this particular combinatorial approach is at best approximative, and other lengths may be suitable as well. The choice of length will depend on such factors as the length and/or complexity of the genome to be detected and compatibility with current nucleic acid amplification and DNA array technologies.

15

In another embodiment, the amplification described above is performed with one of the PCR primers being labelled with biotin, providing means for purification of the labelled products with streptavidin-labelled substrates (12) or other similar methods. The amplified mixture of

20 unlabelled oligonucleotides specific to one template is hybridized with labelled mixtures of oligonucleotides selected for specificity to one or several nucleic acid templates, and the unbound material is collected. In this manner, a mixture of nucleotides which is enriched for nucleic acids present in the unlabelled library only can be generated.

25

The process is based on stringent hybridization. Furthermore, high fidelity hybridization between pools of oligonucleotides and templates (genomic DNA or cDNA) is the basic mode of transfer of genomic

information into OLs. An efficient subtractive hybridization procedure is used to accommodate the features of the aforementioned OLs.

The present invention is illustrated in further detail by the following non-limiting example.

EXAMPLE 1

Generation of OLs, Use Thereof in Subtractive Hybridization to Generate Subtractive Oligonucleotides Libraries (SOLs), and Use of OLs or SOLs in Hybridization Experiments

EXPERIMENTAL METHODS

DNA / oligonucleotides

The starting random DNA pool was synthesised by GIBCO BRL (Burlington, Canada), (RAN), 5'-GCCTGTTGTGAGCCTCCTGTCGAA-N₂₀-TTGAGCGTTTATTCTTGTCTCCC-3'. The corresponding left and right arms were (LEFT) 5'-GCCTGTTGTGAGCCTCCTGTCGAA-3' and (RIGHT) 5'-BioGGGAGACAAGAATAAACGCTCAA-3'. The 5'-end biotinylated oligonucleotides were used to pool out complement strands, using BioMag magnetic particles (PerSeptive Biosystems, Framingham, MA). During preparative hybridization, the left and right arms were blocked by (LEFT) 5'-TTCGACAGGAGGCTCACAAACAGGC-3' and (RIGHT) 5'-GGGAGACAAGAATAAACGCTCAA-3'. These oligonucleotides are termed 'blockers' in the text.

The following genomic DNA was used to produce OL: Adenovirus DNA Type 2, (GIBCO BRL), Lambda DNA cl857 *ind1 Sam 7* (New England Biolabs), pBluescript II SK(+) (Stratagene, San Diego, CA). The Human HeLa DNA used as one control was from Clontech (Palo Alto, CA).

Blotting genomic DNA

The genomic DNA was denatured 2-3 minutes at 95°C and cooled on ice. The nylon membrane (Hybond-N, Amersham Pharmacia Biotech, Piscataway, NJ) was blotted with 100 ng of denatured genomic DNA, dried for 2 minutes on a hot plate and exposed to UV light for 8 minutes. The prehybridization was done for a minimum of 30 minutes in the hybridization buffer (7% SDS, 0.25M Na₂HPO₄ pH7.4, 1mM EDTA, pH 8.0 and 10g/L of BSA).

10 Hybridization and washing of the starting random pool

The preparative hybridization between random core (20N) and targeted DNA was done with 10 pmoles of starting random pool (RAN). The random pool was pre-mixed with 100 pmoles (10 times more than RAN) of LEFT and RIGHT blockers in order to exclude cross-hybridization of left and right arms with genomic DNA. The oligonucleotide mixture was heated up to 95°C, cooled at room temperature and added to the hybridization buffer. The hybridization was done overnight at 50° C. The first washing was done with 6X SSC, followed by subsequent 2X SSC washing at the same temperature as hybridization was done.

20

Generating OL by PCR

The dot containing the genomic DNA and bound probes was cut out of the nylon membrane (radius of 2-4mm), soaked in 100 µl H₂O and heated to 95°C for 1-2 minutes. The solution containing the denatured probe (originally RAN) was then collected and passed through a Sephadex G-50 column in order to eliminate salts and SDS. The PCR was prepared under standard conditions, typical for SELEX-like amplification of DNA (10, 13). The RIGHT 5'-end biotinylated primer of the sense strand (the

25

one which did not hybridise with genomic DNA) and LEFT primer of antisense strand were used in the PCR reaction. The temperature cycles were 53°C, 72°C, 95°C, each 30 seconds, repeated 20 times.

Probe labelling and hybridization

- 5 Before labelling, the PCR reaction mixtures were passed through Sephadex G-50 columns. Around 100-200 ng of PCR product was labelled with 50 pmols of γP^{32} ATP (6000 Ci/mmol, I.C.N. pharmaceuticals, Irvine, CA). The total amount of probe radioactivity was 300 000 c.p.m. The probe was added into 0.5 ml of hybridization buffer. The blotting of genomic
- 10 DNA was done as described above. Hybridizations were done overnight at 50°C. The nylon membrane was washed as previously described, and exposed to Kodak X-OMAT film.

OL labelling and analytical hybridization

- The generated OL was tested, using 1) the original genomic DNA from
- 15 which they were selected (positive control) and 2) using the unrelated genomic DNA (negative control). The OL labelling, hybridization and probe washing was done as described, except that hybridization time was shorter (60 minutes).

Southern blot hybridization

- 20 Electrophoresis was performed in a 1% agarose gel with TBE buffer (80 mM Tris borate, pH 8.0, 2mM Na₂EDTA) and stained with ethidium bromide. One ug of BstEII-digested lambda DNA, 300 ng of adenoviral DNA and 1 µg of AluI-HpaI-digested human HeLa DNA were run on the gel according to specifications (all restriction enzymes used in this work
- 25 were purchased from New England Biolabs). For Southern hybridization, DNA was transferred to Nylon membranes by capillary blot procedure

following manufacturer's recommendations (Amersham Pharmacia Biotech). Hybridization was performed as described above with adenoviral OL. Autoradiographic exposure (using Kodak X-OMAT film) was done at room temperature, for few hours. Stripping of the membrane
5 was done by boiling a 1% SDS solution and pouring it over the Nylon membrane.

Subtractive enrichment of OL

The tester OL (mixed OL) that reflects the two genomes (Adenovirus type
10 2 and Lambda) was made by preparing OL from equimolar mixtures of 2 genomes. The driver OL was produced from the lambda genome only. The production of sense strand (the one which did not hybridize with genomic DNA) was done using 5'-end biotinylated primer in PCR reaction. After denaturing PCR product, the biotinylated sense strand was
15 bound to streptavidin magnetic particles (200 μ g, binding capacity > 200 pmols of biotinylated oligonucleotides, Biomag Magnetic Particles, PerSeptive Biosystems), and pulled-out using a magnet. The complementary antisense strand was discarded with the liquid phase. The mixed antisense tester OL (Lambda + Adenovirus DNA) was
20 produced in the same way. This time, the supernatant with the antisense, non-biotinylated strand was hybridized overnight at 50°C with 10 times molar excess of driver Lambda sense stand attached to magnetic beads. The hybridization buffer was the same as described above but without SDS. After removing the fraction bound to the magnetic beads, the rest
25 of the mixture was used in the analytical hybridization step.

RESULTS

The starting random pool of oligonucleotides contains 4^{20} (i.e. 10^{12}) different 20-mers. The diversity of the sequence motifs is approximately

10¹¹ higher than the diversity of the most complex genomes. A schematic representation of the procedure for generating OL is presented in Figure 1 and is described in detail in the Experimental Methods, above.

Blockers were used in order to avoid hybridization of the flanking arms to the targeted genome, and this step was found to be critical to achieve specificity. The stringency of hybridization conditions eliminates unbound 20-mers, leaving the specific oligonucleotides bound to the membrane via hybridization of the random core to the genome (Fig. 1). This ensemble of selected oligonucleotides constitutes the OL.

10

It should be noted that the starting random pool of oligonucleotides contained about 8 copies of each sequence motif during the first hybridization step (10-20 pmoles) and that the number of copies of each particular 20-mer present in the random mixture was smaller than the number of genome copies.

15

Figure 2 shows that OLs are able to discriminate genomes with complexities around 10³ to 10⁴. The starting random pool of probes binds to all three genomes equally (Fig. 2, row 1). After one round of selection, the OL can hybridise specifically towards a single targeted genome (Fig. 2, rows 2, 3 and 4). The OL can be selected against a mixture of two genomes and the specificity is conserved for both genomes (Fig. 2, row 5).

20

A Southern blot was performed in order to document the distribution of adenovirus OL probes along the genome (Fig. 3). There was no apparent cross-hybridization of adenovirus OL to either HeLa or Lambda DNA (Fig. 3b, lanes 1, 4, 5 and 6). The intensity of radioactive signal over adenoviral genome generated by adenovirus-specific OL was linearly increasing with

25

the DNA fragments' length (Fig. 4). Therefore, one could deduce a uniform distribution of OL throughout the genomic DNA.

In the next step, only the subset of the adenovirus OL bound to the 3648
5 bp band in Figure 3b (rows 3 and 6) was reamplified and selected. The membrane was washed from the original probe and hybridised with the 3648 bp subset of the original OL. Figure 3c shows that the specificity of the OL subset is obtained against the 3648 bp band. These data illustrate the successful increase in the specificity and the reduction in the
10 complexity of the original Adeno-specific OL to that of the 3648 bp subset, using just one additional round of selection.

One round of subtractive enrichment between two oligonucleotide libraries was performed as schematized (Fig. 5a). The tester OL reflects
15 the two genomes (adenovirus type 2 and lambda phage). The driver OL was produced from the Lambda genome only. The single stranded (ss) OL from the driver DNA was used to pool out the complementary single stranded, mixed, tester OL. After removing the subtracted fraction, the rest of ssDNA was used as a probe in the analytical hybridization step.
20 The intensity of hybridization signals between Lambda and Adeno genomes, before (Fig. 5b) and after (Figure 5c) one round of subtractive enrichment was shown. It should be noted that further subtraction steps could be performed by changing the sequence design of flanking arms between tester and driver OLs, as suggested by recent developments in
25 subtractive procedures (14).

With reference to Figure 6, the relative distribution of 20-mers with different numbers of mismatches that hybridized to the targeted DNA was predicted. The number of 20-mers (N) with (m) number of mismatches

($m=0, 1, \dots, 20$) capable of hybridizing to the target sequence were then calculated. First, the number of combinations of 20-mers (C) with the same number of mismatches (m) in the initial random pool of oligonucleotides that are capable of hybridizing to a specific 20-mer motif was calculated. Since each full match could be replaced by 3 different mismatches, the number of combinations must be multiplied by 3^m i.e. $C \cdot 3^m$. Finally, these numbers were adjusted to reflect the sequence-dependent thermostability of 20-mers with ($m = 0, 1 \dots 20$) number of mismatches which hybridised at 52°C using the thermostable fraction of the binomial distribution for each n-mer population. Duplexes of more than 7 mismatches are not observed based on this thermostability criterion. Therefore, the majority of 20-mers captured after the first round of selection will harbour less than 6 mismatches.

A process that generates amplifiable DNA oligonucleotide libraries which are specific for a given segment of DNA has been described. This process is akin to random priming, because it is possible to generate probes without *a priori* knowledge of the template sequence. One round of preparative hybridization was enough to produce genome-specific oligonucleotide libraries (Figs. 2 and 3). The OLs were inferred from genomes of complexity of 10^3 - 10^4 .

The process described herein generates probes with high detection power. These probes/selected oligonucleotides can contain mismatches. The notion that introduction of artificial mismatches could increase detection power of oligonucleotides during single nucleotide polymorphism (SNP) detection was well documented by Guo *et al* (6). However, the prediction of positions and types of mismatches, which should be introduced to increase detectability of oligonucleotide, remains

undefined. Consequently, to enhance oligonucleotide detectability by introducing (artificial) mismatches, one must search different positions and types of mismatches along the oligonucleotide. Once they are empirically determined, i.e. tested on 2 different sequence motifs, the
5 oligonucleotide containing particular mismatches could be used (15).

The present process provides an approach based on differential selection of thermostable oligonucleotides (i.e. their differential stability), which are present in one, but not in the second system. The selection of oligonucleotides with the highest detectability is inherently present in this
10 process, i.e. the method suggests a solution to the problem of where and what type of mismatches should be introduced to increase detection power of oligonucleotide, or to find the particular oligonucleotide which best discriminates between 2 sequence motifs which may differ by a single base.

15

Without wishing to be bound by any hypothesis, the following provides an explanation of what is believed to be occurring during the process of the present invention. Based on calculations, it is expected that the 20-mers selected in an OL can contain up to 6 mismatches. Nevertheless,
20 specificity toward a given template was achieved, suggesting that the presence of these putative mismatches did not interfere with good discrimination. In other words, mismatch-free hybridization is not critical for differential detection approach; rather, the *relative differences* in the thermodynamical stabilities of the hybridized oligonucleotides appear to
25 be determinative. The present process uses selection of oligonucleotides based on this criterion and therefore provides the possibility of overcoming current technological limitations. In the second and further rounds of selection, the number of 20-mers both in the targeted genome and the probe mixture (OL) could be adjusted. Each new round of

preparative hybridization (Fig.1) and/or subtraction (Fig. 6) could reduce the complexity of OL, by using the excess OL rather than genomic DNA. Therefore, the average number of mismatches for each particular 20-mer will continue to decline until it reaches the sequence-dependent limitation, but not the concentration-dependent limitation.

In summary, OLs are generated from the template DNAs. These OLs are used in subtractive hybridization, for example between genomic or cDNA-based libraries (OL1 and OL2) to make a new Subtractive Oligonucleotide Library (SOL1/2 and/or SOL2/1)), that is/are specific for one system/library but not for the other. Oligonucleotides isolated from such subtractive libraries (SOL) are useful for diagnostic purposes. They can a) directly serve as highly specific hybridization probes or b) they can be tested for PCR-specific differential amplification, specific for one, but not the other biological system.

These libraries (OL or SOL) can be hybridized to oligonucleotide chip arrays in order to obtain a specific hybridization pattern that is useful for diagnostic features: each OL produces an image which is specific for the templated DNA (genome or cDNA). A particular advantage in using OL or SOL instead of genomic/cDNA libraries is that the hybridization signal is not dependent on copy number and distribution of particular sequence motifs. By comparing images of different genomes/cDNA, one can deduce which oligonucleotides are highly specific for a single genome/cDNA, and use this or these oligonucleotide(s) as "genome tags". The oligonucleotides obtained can also be used for specific diagnostic PCR.

OLs or SOLs can be inferred from two biologically relevant systems, like mammalian cells, to detect fine differences in cell cycle, tissue status, viral infection, age/development status etc.

- 5 Although the present invention has been described hereinabove by way of a preferred embodiment, it can be modified by one of skill in the art without departing from the spirit and nature of the subject invention, as defined more particularly in the appended claims.

LIST OF REFERENCES

1. Diatchenko, L., Lukyanov, S., Lau, Y.F. and Siebert, P.D.
Suppression subtractive hybridization: a versatile method for
5 identifying differentially expressed genes. *Methods Enzymol.*,
1999, **303**, 349-380.
2. Nilsson, P., Larsson, A., Lundeberg, J., Uhlen, M. and Nygren,
P.A. (1999) Mutation scanning of PCR products by subtractive
10 oligonucleotide hybridization analysis. *Biotechniques*, **2**, 308-
315.
3. Whitcombe, D., Newton C.R. and Little, S. (1998) *Curr. Opin.*
Biotechnol., **9**, 602-608.
- 15 4. Watson, A., Mazumder, A., Stewart, M. and Balasubramanian, S.
(1998) *Curr. Opin. Biotechnol.*, **9**, 609-614.
5. Fredricks D.N. and Relman D.A. (1999) Application of
20 polymerase chain reaction to the diagnosis of infectious
diseases. *Clin Infect Dis.*, **29**, 475-86;
6. Gerhold, D., Rushmore, T. and Caskey, C.T. (1999) *Trends*
Biochem. Sci., **24**, 168-73.
- 25 7. Matz M.V. and Lukyanov S.A. (1998) Different strategies of
differential display: areas of application. *Nucleic Acids Res.*, **26**,
5537-43.
- 30 8. Guo, Z., Qiunghua, L. and Smith, L.M. (1997) *Nat. Biotechnology*,
15, 331-335.
9. Akopyants *et al*, *Proc. Natl. Acad. Sci. USA* 95:13108-13113.
- 35 10. U.S. Patent No. 5,270,163 (Gold *et al*), December 14, 1993:
Methods for Identifying Nucleic Acid Ligands.
- 40 11. Armour, J.A., Sismani, C., Patsalis, P.C. and Cross, G. (2000)
Nucleic Acids Res., **28**, 605-609.

12. Mitchell L.G. and Merril C.R. (1989) Affinity generation of single-stranded DNA for dideoxy sequencing following the polymerase chain reaction. *Anal Biochem.*, **178**, 239-42.
- 5 13. Fitzwater, T. and Polsky, B. (1996) *Methods Enzymol.*, **267**, 275-301. Combinatorial Chemistry, A Selex Primer. Academic Press, New York.
- 10 14. Diatchenko L., Lau Y.F., Campbell A.P., Chenchik A., Moqadam F., Huang B., Lukyanov S., Lukyanov K., Gurskaya N., Sverdlov E.D. and Siebert P.D. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci.*, **93**, 6025-30; Luo, J.-H. *et al.*, Suppression subtractive hybridization: a method for
15 generating differentially regulated or tissue-specific cDNA probes
and libraries. *Nucleic Acids Res.*, 1999, **27**, e24.
- 20 15. Maldonado-Rodrigues, R., Espinosa-Lara, M., Loyola-Abita, P., Beattie, W.G. and Beattie, K.L. (1999) *Molecular Biotechnology*, **1**, 13-25; Okano, K., Uematsu, C., Matsunaga, H. and Kambara, H. (1998), *Electrophoresis*, **18**, 3071-3078.

WHAT IS CLAIMED IS:

1. A process for generating an oligonucleotide library which originates from a chosen biological material, comprising:
 - 5 a) generating random oligonucleotides, wherein said oligonucleotides are of a uniform length comprising a central segment of randomly varied bases and segments of a defined sequence flanking the central segment on each side;
 - b) hybridizing said random oligonucleotide mixture of a) with a nucleic acid
 - 10 template of biological origin under hybridization conditions which enable the formation of duplexes, while minimizing or abrogating mismatches;
 - c) separating said duplexes from non-duplexed material; and
 - d) amplifying said hybridized oligonucleotides.
- 15 2. The process of claim 1, wherein said template is genomic DNA or cDNA.
3. The process of claim 1 or 2, wherein said biological material is selected from genomic DNA and cDNA.
- 20 4. A method of subtracting between different oligonucleotide libraries which contain common sequence motifs comprising:
 - a) partitioning double stranded OL under denaturing conditions to separate: (i) antisense strands hybridized with template and (ii) sense
 - 25 strands;
 - b) annealing said antisense strand of OL1 with an excess of OL2 sense strand, under hybridization conditions:

c) partitioning double stranded hybrids (OL1 antisense/ OL2 sense) and single stranded OL2 sense molecules from single stranded antisense OL1 molecules;

d) amplifying OL1 antisense molecules; and

- 5 e) repeating steps a) to d) until OL1 antisense does not have complementary sequences to OL2 sense.

5. The method of claim 4, wherein said partitioning is carried out using streptavidin/biotin.

10

6. A process for generating an oligonucleotide library which originates from a chosen biological material, comprising:

a) generating a random oligonucleotide mixture, wherein said oligonucleotides are of a uniform length;

- 15 b) hybridizing said random oligonucleotide mixture of a) with a nucleic acid template of biological origin under hybridization conditions which enable the formation of duplexes, while minimizing or abrogating mismatches;

c) separating said duplexes from non-duplexed material;

d) amplifying said hybridized oligonucleotides; and

- 20 e) subtracting between different oligonucleotide libraries which contain common sequence motifs.

7. The process of claim 6, wherein said subtracting between different oligonucleotide libraries which contain common sequence motifs

25 comprises:

a) partitioning double stranded OL under denaturing conditions to separate: (i) antisense strands hybridized with template and (ii) sense strands;

- b) annealing said antisense strand of OL1 with an excess of OL2 sense strand, under hybridization conditions;
 - c) partitioning double stranded hybrids (OL1 antisense/ OL2 sense) and single stranded OL2 sense molecules from single stranded antisense OL1 molecules;
 - d) amplifying OL1 antisense molecules; and
 - e) repeating steps a) to d) until OL1 antisense does not have complementary sequences to OL2 sense.
8. The process of claim 7, wherein said partitioning is carried out using streptavidin/biotin.
9. The process of any one of claims 6 to 8, wherein said template is genomic DNA or cDNA.
10. The process of any one of claims 6 to 9, wherein said biological material is selected from genomic DNA and cDNA.
11. An oligonucleotide library produced by the process of any one of claims 1, 2, 3, 6, 7, 8, 9 or 10.
12. Use of the oligonucleotide library of claim 11 in a diagnostic kit.
13. Method of diagnosis comprising use of the oligonucleotide library of claim 11.
14. Use of the oligonucleotide library of claim 11 in reversed dot blots wherein said oligonucleotide library is covalently bound to membranes.

15. Use of the oligonucleotide library in claim 11, wherein said oligonucleotide library is hybridized against oligonucleotide chip arrays.

5

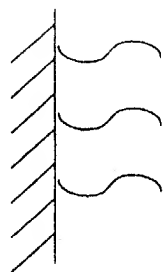
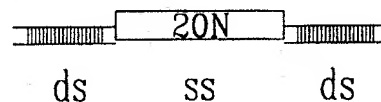
10

15

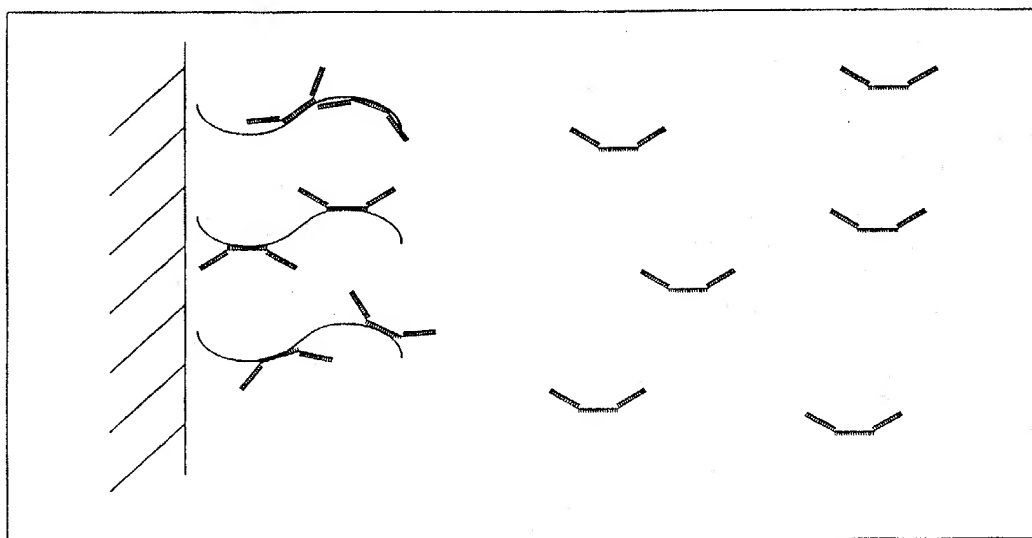
20

1/7

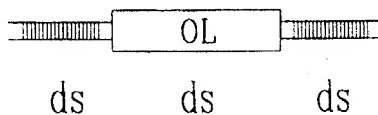
Membrane-bound denatured target DNA

Random 20-mer core with
left and right blockers

Preparative hybridization


















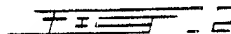
1. Wash unbound OL
2. Elute bound OL
3. PCR amplify bound OL



SUBSTITUTE SHEET (RULE 26)

2/7

OL Probes		Genomes		
		Adenovirus	pBluescript	Lambda
Random	OL			
Adenovirus	OL			
pBluescript	OL			
Lambda	OL			
Adenovirus Lambda mixed	OL			



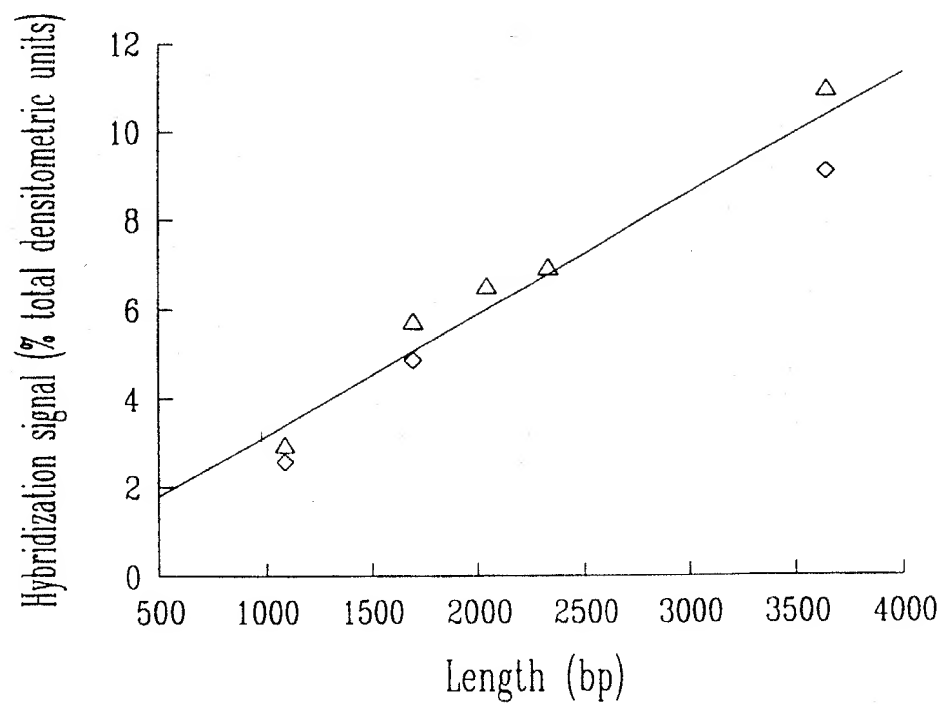
SUBSTITUTE SHEET (RULE 26)

3/7

[illegible]

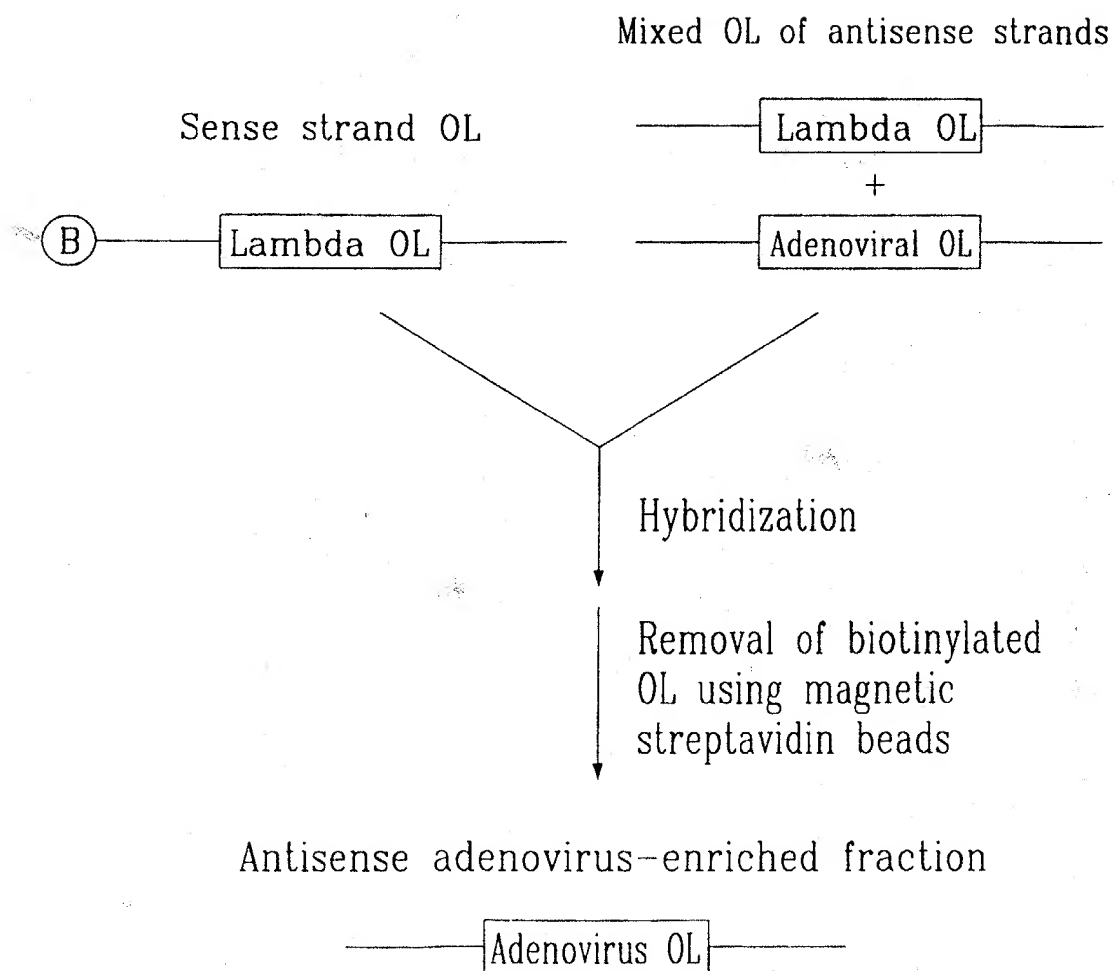
SUBSTITUTE SHEET (RULE 26)

4/7

FIG. 4

SUBSTITUTE SHEET (RULE 26)

5/7

FIG. 5A

6/7

Adenovirus Lambda



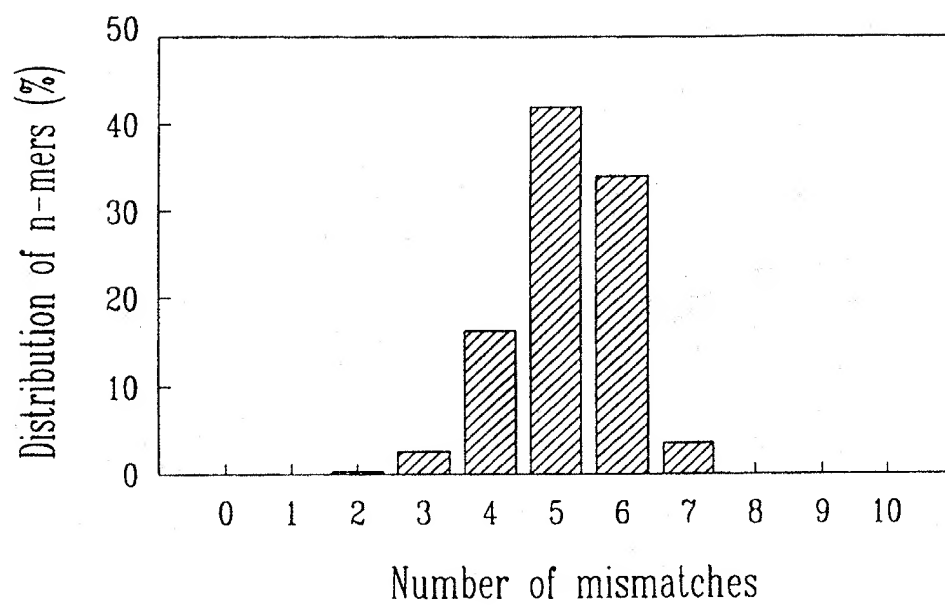
FILE - 5C

Adenovirus Lambda



FILE - 5B

7/7

FIG. 6

INTERNATIONAL SEARCH REPORT

Int. Application No
PCT/CA 00/00047

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 98 15651 A (BRAX GENOMICS LTD ; THOMPSON ANDREW HUGIN (GB); SCHMIDT GUNTER (GB)) 16 April 1998 (1998-04-16) the whole document	1-15
Y	WO 98 15649 A (CREMER THOMAS ; CRAIG JEFF (DE); UNIV HEIDELBERG (DE)) 16 April 1998 (1998-04-16) the whole document	1-15
Y	SEDLACEK ET AL.: "Direct selection of DNA sequences conserved between species" NUCLEIC ACIDS RESEARCH, vol. 21, no. 15, 1993, pages 3419-3425, XP002138409 the whole document	1-15
	-/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

23 May 2000

Date of mailing of the international search report

09/06/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel: (+31-70) 340-2040, Tx: 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Hagenmaier, S

INTERNATIONAL SEARCH REPORT

International Application No

PCT/CA 00/00047

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 98 26098 A (FELDHAUS MICHAEL JOHN ;KAMB ALEXANDER (US); VENTANA GENETICS INC () 18 June 1998 (1998-06-18) the whole document ---	1-15
Y	GENG ET AL.: "ISOLATION OF DIFFERENTIALLY EXPRESSED GENES BY COMBINING REPRESENTATIONAL DIFFERENCE ANALYSIS (RDA) AND CDNA LIBRARY ARRAYS" BIOTECHNIQUES,US,EATON PUBLISHING, NATICK, vol. 25, no. 3, 5 September 1998 (1998-09-05), pages 434-438, XP002916349 ISSN: 0736-6205 the whole document ---	1-15
Y	WO 97 27317 A (CHEE MARK ;LAI CHAOQIANG (US); LEE DANNY (US); AFFYMETRIX INC (US)) 31 July 1997 (1997-07-31) See "Modifying sample to improve signal to noise ration", page 43-page 45 the whole document ---	1-15
A	WO 97 13877 A (LYNX THERAPEUTICS INC ;MARTIN DAVID W (US)) 17 April 1997 (1997-04-17) the whole document ---	
A	WO 96 34875 A (MORRIS KEVIN N ;NEXSTAR PHARMACEUTICALS INC (US); CHEN HANG (US);) 7 November 1996 (1996-11-07) the whole document ---	
A	WO 98 42871 A (BOEHRINGER MANNHEIM CORP) 1 October 1998 (1998-10-01) the whole document ---	
A	WO 91 19813 A (UNIV COLORADO FOUNDATION) 26 December 1991 (1991-12-26) the whole document ---	
A	WO 93 04204 A (ISIS PHARMACEUTICALS INC) 4 March 1993 (1993-03-04) the whole document ---	
P,X	WO 99 28497 A (BELMAAZA ABDELLAH ;BRUKNER IVAN (CA); POLYGENE INC (CA); ROUISSI N) 10 June 1999 (1999-06-10) the whole document ---	1-15
	-/--	

INTERNATIONAL SEARCH REPORT

Inventor's Application No
PCT/CA 00/00047

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>ARMOUR ET AL.: "MEASUREMENT OF LOCUS COPY NUMBER BY HYBRIDISATION WITH AMPLIFIABLE PROBES"</p> <p>NUCLEIC ACIDS RESEARCH,</p> <p>vol. 28, no. 2, 2000, pages 605-609,</p> <p>XP002138423</p> <p>cited in the application</p> <p>the whole document</p> <p>-----</p>	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/CA 00/00047

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9815651 A	16-04-1998	AU 4565297 A CN 1232510 A EP 0956366 A	05-05-1998 20-10-1999 17-11-1999
WO 9815649 A	16-04-1998	AU 4116397 A EP 0951566 A	05-05-1998 27-10-1999
WO 9826098 A	18-06-1998	AU 5605798 A EP 0948649 A	03-07-1998 13-10-1999
WO 9727317 A	31-07-1997	AU 2253397 A EP 0880598 A	20-08-1997 02-12-1998
WO 9713877 A	17-04-1997	AU 712929 B AU 4277896 A AU 6102096 A AU 7717596 A CN 1193357 A CZ 9700866 A CZ 9703926 A EP 0793718 A EP 0832287 A EP 0931165 A FI 971473 A HU 9900910 A JP 11507528 T JP 10507357 T NO 971644 A NO 975744 A PL 324000 A WO 9641011 A	18-11-1999 06-05-1996 30-12-1996 30-04-1997 16-09-1998 17-09-1997 17-06-1998 10-09-1997 01-04-1998 28-07-1999 04-06-1997 28-07-1999 06-07-1999 21-07-1998 02-06-1997 05-02-1998 27-04-1998 19-12-1996
WO 9634875 A	07-11-1996	US 5789157 A US 5712375 A US 5763566 A US 5864026 A AU 5670796 A CA 2219807 A EP 0823914 A US 6013443 A JP 11505106 T	04-08-1998 27-01-1998 09-06-1998 26-01-1999 21-11-1996 07-11-1996 18-02-1998 11-01-2000 18-05-1999
WO 9842871 A	01-10-1998	US 5958738 A EP 0920536 A	28-09-1999 09-06-1999
WO 9119813 A	26-12-1991	AT 160821 T AU 663053 B AU 8206191 A CA 2084987 A DE 69128350 D DE 69128350 T DK 533838 T EP 0533838 A EP 0786469 A ES 2110444 T GR 3025594 T IL 98456 A IL 112141 A	15-12-1997 28-09-1995 07-01-1992 12-12-1991 15-01-1998 26-03-1998 23-02-1998 31-03-1993 30-07-1997 16-02-1998 31-03-1998 18-03-1997 05-04-1998

Form PCT/ISA/210 (patent family annex) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Application No

PCT/CA 00/00047

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9119813 A		JP 8252100 A	01-10-1996
		JP 2763958 B	11-06-1998
		KR 9702255 B	26-02-1997
		NZ 238492 A	27-07-1993
		US 5459015 A	17-10-1995
		US 5707796 A	13-01-1998
		US 5472841 A	05-12-1995
		US 5543293 A	06-08-1996
		US 5849479 A	15-12-1998
		US 5683867 A	04-11-1997
		US 5503978 A	02-04-1996
		US 5527894 A	18-06-1996
		US 5650275 A	22-07-1997
		US 5637459 A	10-06-1997
		US 5648214 A	15-07-1997
		US 5723289 A	03-03-1998
		US 5629155 A	13-05-1997
		US 5756287 A	26-05-1998
		US 5641629 A	24-06-1997
		US 5639868 A	17-06-1997
		US 5622828 A	22-04-1997
		US 5705337 A	06-01-1998
		US 5637461 A	10-06-1997
		US 5696249 A	09-12-1997
		US 5670637 A	23-09-1997
		US 5660985 A	26-08-1997
		US 5750342 A	12-05-1998
		US 5688935 A	18-11-1997
		US 5763566 A	09-06-1998
		US 5712375 A	27-01-1998
		US 5789157 A	04-08-1998
		US 6011020 A	04-01-2000
		US 5864026 A	26-01-1999
		US 5637682 A	10-06-1997
		US 5734034 A	31-03-1998
		US 5595877 A	21-01-1997
WO 9304204 A	04-03-1993	AU 2547592 A	16-03-1993
		EP 0673435 A	27-09-1995
		JP 7501929 T	02-03-1995
		US 5672472 A	30-09-1997
		US 5698391 A	16-12-1997
WO 9928497 A	10-06-1999	US 5747253 A	05-05-1998
		AU 1477299 A	16-06-1999